



Competence Center - Big Data

2nd Informal Meeting - 14th July 2017

ATTENDEES (via Vidyo)

Alexandre Lindote, Ana Peixoto, Ana Sofia Nunes, António Esteves, Bruno Galhardo, Celso Franco, Claudio Silva, David Fernandes, Filipe Veloso, Giles Strong, Guilherme Milhano, Helmut Wolters, Henrique Carvalho, Juan Pedro Araque, Liliana Apolinário, Lorenzo Cazon, Luís Silva, Marcin Stolarski, Nuno Castro, Ricardo Gonçalo, Tiago Vale, Vladimir Solovov

NOTES

- **Agenda:** <https://indico.lip.pt/indico/conferenceDisplay.py?confId=275>
- Survey on the tools being used and interesting to be used:

Name	Tools being used	Tools which could be used in the context of your ongoing projects
Nuno Castro	TMVA (ANN, BDT)	Machine learning: TMVA (k-fold cross-validation), Keras, TensorFlow, Theano, SK-Learn, GPU training, CNNs Data handling: GlusterFS
Giles Strong	Theano, Keras (classification & regression), SK-Learn (pre-processing, k-fold CV, PCA), XGBoost (Classification), Pandas	Tensorflow GPU training Distributed training RNNs & CNNs
António Esteves	Theano, Keras	Theano, Keras, TensorFlow, CNNs and RNNs, acceleration of computation with GPUs and FPGAs
Alexandre Lindote	NN, computation with GPUs	BDT, SVM, kNN, feature detection algorithms (supervised and unsupervised)
Tiago Vale	TensorFlow	Neural Nets, Random Forests
Luís Silva	ANN, Scikit Learn (preprocessing, k-fold, CV and supervised and unsupervised techniques), numpy, pandas, Octave/Matlab.	TensorFlow, Theano, Keras, working with GPUs.
Celso Franco	Dynamic Neural Network with backpropagation algorithm, Recurrent Neural Network, SOM in Matlab and Octave, SVM	Theano, Keras, TensorFlow

- Looks clear that we have a significant expertise, but also that there is interest in a training program.
 - Significant overlap in the interests. Deep learning looks the natural candidate to get started with.
 - Important to use the survey as a live document, so we should continue to update it:
https://docs.google.com/document/d/1Cd0vLKFSb8860uMb583Vo9YPdSG_8JL9q8N0HHPaN4/edit
 - Would be good to have a few computers - possibly with a GPU board - with all the required ML software to have a suitable testbed for training and development. Possibilities will be discussed with the Computing Infrastructure team.
- **Lorenzo Cazon**, *Discussion on the proposal of a School & Workshop on Data Science in (Astro)Particle Physics and the bridge to industry.*
 - Proposal looks very promising.
 - Organizing committee being constituted: please get in touch with Lorenzo if you would like to contribute.
 - Aiming to Spring 2018, in Lisbon. Date to be settled before Summer.
 - Urgent to contact speakers and companies.
 - **Giles Strong**, *tutorial on Keras.*
 - https://github.com/GilesStrong/ML_Tutorials
 - Giles is available to help with follow-up questions.

NEXT MEETING

Next meeting will be organized in early October. Doodle poll to be circulated early in September. Please volunteer for the journal club during the next sessions.